NOISE REDUCTION METHOD

BACKGROUND OF THE INVENTION

1. Field of the Invention

The present invention relates to a noise reduction method and, more particularly, to a method using spectral subtraction to reduce noise.

2. Description of Related Art

The spectral subtraction method has been proven effective in enhancing speech degraded by additive noise. It is simple to implement, hence is suitable as the pre-processing scheme for speech coding and recognition applications. This method subtracts the noise spectrum estimate from the noisy speech spectrum to estimate the speech magnitude spectrum, so as to obtain the clean speech signals.

FIG. 1 shows the flowchart of the aforementioned spectral subtraction method, wherein the input noisy speech is divided into a plurality of continuous frames, and each frame is represented by an additive noise model :

$$y_r(k) = s_r(k) + w_r(k),$$

where $y_r(k)$, $s_r(k)$ and $w_r(k)$ denote respectively the k-th noisy speech, clean speech, and noise sample of the r-th frame. Taking the fast Fourier transform of the noisy speech frame $y_r(k)$ (step S101), the noisy speech spectrum of the r-th frame at the k-th frequency component is obtained and denoted as $|Y_r(k)|^2$. In addition, the noisy speech $y_r(k)$ is also applied in a silence detection process (step S102) and a noise spectrum estimation process (step S103) to estimate a noise spectrum, denoted as $|W_r(k)|^2$. After performing a spectral subtraction process (step S104), the energy spectrum of clean

speech is obtained as follows:

$$|\hat{S}_r(k)|^2 = |Y_r(k)|^2 - |W_r(k)|^2. \tag{1}$$

If the phase spectrum of the clean speech can be approximated by the phase spectrum of the noisy speech, the estimate of clean speech $\hat{s}_r(k)$ can be obtained by taking the inverse fast Fourier transform of $|\hat{S}_r(k)|^2$.

Such a method is suitable as the pre-processing scheme for speech coding and recognition applications because it is easy, effective and simple to implement. However, the noise spectrum estimate may cause a relatively large spectral excursion in the spectrum estimate of clean speech. This spectral excursion will be perceived as time varying tones contributing to the so-called musical noise.

To reduce the musical noise Berouti et al proposed a noise reduction method to over-subtract the noise spectrum estimate, and a description of such can be found in M. Berouti, R. Schwartz, and J. Makhoul "Enhancement of speech corrupted by acoustic noise", pp.208-211, 1979 IEEE, which is incorporated herein for reference, wherein the formula (1) is modified as:

$$|\hat{S}_r(k)|^2 = |Y_r(k)|^2 - \alpha_r \cdot |W_r(k)|^2. \quad \alpha_r \geqq 1, \tag{2}$$

so as to decrease the influence caused by the excursion of the noise spectrum estimate and thus reduce the effect of musical noise. In the method, the over-subtraction factor α_r was determined by the signal-to-noise ratio (SNR) of the processing frame, and can be expressed by formula:

$$\alpha_r = \alpha_0 + \mathrm{SNR}_r \cdot \frac{1-\alpha_0}{\mathrm{SNR}_1}, \tag{3}$$

where α_0 is pre-selected over-subtraction factor when SNR =0, SNR_1 is pre-selected SNR value when α_r=1, SNR_r is the estimate of signal-to-noise ratio of the processed r-th frame. Based on the formula (3), it is known that α_r is inversely proportional to SNR_r. The smaller the SNR_r is, the larger the α_r is, and a larger α_r is helpful in removing the larger noise spectrum excursion.

Examining human speech spectrum, it is known that the speech energy distributes non-uniformly and often concentrates on lower frequency components. Hence SNR differs with frequencies and often have larger values at lower frequency components. From the formula (3), it is known that more suppression is needed for lower SNR and vise versa. High-frequency components thus need more suppression to avoid musical noise, while low-frequency components need less suppression to prevent speech distortion. However, for the over-subtraction method based on formulas (2) and (3), it faces the problem of too much over-subtraction and hence speech distortion at low-frequency components while too less over-subtraction and hence musical noise at high-frequency components. Accordingly, improved schemes are proposed to avoid such a problem, and one of the schemes can be found in Kuo-Guan Wu and Po-Cheng Chen "Efficient speech enhancement using spectral subtraction for car hands-free application". 2001 Digest of technical papers, pp. 220-221, which is incorporated herein for reference. However, it is unable to completely eliminate the problem. Therefore, there is a need for the above conventional noise reduction method to be improved.

SUMMARY OF THE INVENTION

The object of the present invention is to provide a noise reduction method capable of effectively eliminating the musical noise and reducing speech distortion.

To achieve the object, the noise reduction method divides input noise speech into a plurality of continuous frames, determines noisy speech spectrum for each frame, and partitions frequency band into multiple sub-bands to determine clean speech spectrum from the noisy speech spectrum on each sub-band. The method is provided to first estimate noise spectrum of r-th frame at k-th frequency component from the noisy speech of r-th frame by silence detection and noise spectrum estimation. Next, the signal-to-noise ratio (SNR) value of i-th sub-band for r-th frame is estimated. Then, an over-subtraction factor of sub-band i is determined based on the estimated sub-band SNR. Finally, the clean speech spectrum estimate is determined by performing a spectral subtraction on each sub-band.

Other objects, advantages, and novel features of the invention will become more apparent from the following detailed description when taken in conjunction with the accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 is the flowchart of a conventional spectral subtraction method.

FIG. 2 is the flowchart of the noise reduction method in accordance with the present invention.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

With reference to FIG. 2, there is shown the flowchart of a preferred embodiment of the noise reduction method in accordance with the present invention. As shown, the input noisy speech of the r-th frame

$y_r(k) = s_r(k) + w_r(k)$ is processes by FFT (fast Fourier Transform) (step S201) to obtain its energy spectrum $|Y_r(k)|^2$. The noisy speech $y_r(k)$ is also processed by silence detection (step S202) and noise spectrum estimation (step S203) to estimate the noise spectrum of the r-th frame, denoted as $|W_r(k)|^2$.

For the noisy speech spectrum $|Y_r(k)|^2$ and noise spectrum $|W_r(k)|^2$, the method of the present invention utilizes a sub-band over-subtraction mechanism to determine the estimate of clean speech spectrum $|\hat{S}_r(k)|^2$, which is then processed by IFFT (Inverse Fast Fourier Transform) (step S207) for being restored to enhanced frame signal $\hat{s}_r(k)$. The method of the present invention partitions the frequency band into multiple sub-bands and perform over-subtraction on each sub-band. To implement over-subtraction on each sub-band, it is first performed a sub-band SNR estimation (step S204) to estimate a SNR value for determining the over-subtraction factor of the sub-band. The SNR value can be obtained by a regression formula as follows:

$$\mathrm{SNR}_r(i) = \mu \cdot \mathrm{SNR}_{r-1}^{o}(i) + (1-\mu) \cdot 10 \cdot \log_{10}\left(\frac{\sum\limits_{k \in \text{sub-band } i} |Y_r(i,k)|^2}{\sum\limits_{k \in \text{sub-band } i} |W_r(i,k)|^2} - 1 \right),$$

where i is the index of sub-band, $\mathrm{SNR}_r(i)$ is SNR estimate of the i-th sub-band for the r-th frame, $|Y_r(i,k)|^2$ is noisy speech spectrum of the *r*-th frame at the k-the frequency component of the i-th sub-band, $|W_r(i,k)|^2$ is the corresponding noise spectrum, $0 < \mu < 1$, and $\mathrm{SNR}_{r-1}^{o}(i)$ is the SNR of the sub-band for the previous frame after noise reduction, which is expressed by the following formula:

$$\mathrm{SNR}^{o}_{r-1}(i) = 10 \cdot \log_{10} \frac{\sum_{k \in \text{sub-band } i} \left|\hat{S}_{r-1}(i,k)\right|^2}{\sum_{k \in \text{sub-band } i} \left|W_{r-1}(i,k)\right|^2},$$

where $\left|\hat{S}_{r-1}(k,i)\right|^2$ is the estimate of clean speech spectrum of the previous, i.e., the (r-1)-th, frame after being processed in the sub-band i.

In step S205, the sub-band over-subtraction factor $\alpha_r(i)$ is determined based on the estimated sub-band SNR value $\mathrm{SNR}_r(i)$, and is expressed by the formula as follows:

$$\alpha_r(i) = \alpha_0(i) + \mathrm{SNR}_r(i) \cdot \frac{1-\alpha_0(i)}{\mathrm{SNR}_1(i)},$$

where $\alpha_0(i)$ is pre-selected over-subtraction factor when the actual $\mathrm{SNR}_r(i)=0$ at sub-band i, and $\mathrm{SNR}_1(i)$ represents pre-selected SNR value when $\alpha_r(i)=1$.

Once determining the over-subtraction factor $\alpha_r(i)$ for each sub-band i, it is able to perform spectral over-subtraction on each sub-band i (step S206), as expressed by the following formula:

$$\left|\hat{S}_r(i,k)\right|^2 = \left|Y_r(i,k)\right|^2 - \alpha_r(i) \cdot \left|W_r(i,k)\right|^2,$$

wherein the determined $\left|\hat{S}_r(i,k)\right|^2$ is the clean speech spectrum at sub-band i for the r-th frame. After performing over-subtraction for each sub-band i, the IFFT is applied (step S207) to obtain the estimated enhanced frame signal $\hat{s}_r(k)$.

In executing the aforementioned method, due to the small number of frequency samples in the lower bands, there will be large variation in sub-band SNR estimate when the noise is strong, which may cause an error in $\alpha_r(i)$ and

influence the quality of the restored speech. To avoid such a problem, in step S205, the SNR value SNR_r of the whole frame is incorporated into modification of sub-band over-subtraction factors as follows:

$$\alpha_r(i) = \alpha_{max} \quad \text{if} \quad SNR_r < SNR_{min},$$

where SNR_{min} is pre-selected minimum value of SNR.

Furthermore, in this embodiment, the step S204 employs regression scheme to estimate the SNR value for determining the over-subtraction factor of the sub-band. However, in practical application, the SNR value of sub-band can also be determined by other known speech signal SNR estimation methods, for example, the high order statistic method described in Elias Nemer, Rafik Goubran and Samy Mahmoud: 'SNR estimation of speech signals using subbands and fourth-order statistics', IEEE Signal Processing Letters, 1999, vol. 6, no. 7, pp. 171-174, which is incorporated herein for reference.

To verify the effect of the present noise reduction method, noisy speech data is generated by adding clean speech data with white Gaussian noise of variant magnitudes to form 3 segmental SNRs: 15dB, 10dB and 5dB. Eight clean speech sentences are collected with 5 sentences from males and 3 from females. Table 1 compares the averaged segmental SNR improvements of conventional over-subtraction method (with parameters of $\alpha_0 = 7.5$ and $SNR_1 = 20$) and those of the present method (with parameters of $\alpha_0(1 \sim 18) = 2$, $SNR_1(1 \sim 13) = 1.5$, $SNR_1(14 \sim 18) = 1.25$) with sub-band SNR obtained from clean speech data.

Table1

Method / Input SNR	Conventional over-subtraction	Present sub-band over-subtraction	Improvement of the present method
15dB	2.39	3.33	39.3%
10dB	3.86	4.76	23.3%
5dB	5.64	6.64	17.5%

From this comparison, it is known that at 15dB input SNR, the present method has the potential of achieving 40% improvement over the conventional method. The potential improvements increase with input SNR.

Table 2 compares the averaged segmental SNR improvements of conventional over-subtraction method (with parameters of $\alpha_0 = 7.5$ and $SNR_1 = 20$) and those of the present method (with parameters of $\alpha_0(1 \sim 18) = 2$, $\mu = 0.25$, $SNR_1(1 \sim 9) = 10$, $SNR_1(10 \sim 13) = 15$, $SNR_1(14 \sim 16) = 2$, and $SNR_1(17 \sim 18) = 1.25$) with sub-band SNR obtained from the step S204 of sub-band SNR estimation.

Table 2

Method / Input SNR	Conventional over-subtraction	Present sub-band over-subtraction	Improvement of the present method
15dB	2.39	2.80	17.0%
10dB	3.86	4.09	6.0%
5dB	5.64	5.96	5.7%

From Table 2, it is known that at input SNR=15dB, although the SNR value of sub-band is obtained by estimation, the present method still can achieve 17% improvement over the conventional method.

Although the present invention has been explained in relation to its preferred embodiment, it is to be understood that many other possible

modifications and variations can be made without departing from the spirit and scope of the invention as hereinafter claimed.